Gmacs: A size-structured stock assessment modeling framework

Presentation to CPT: September 15th 2014
by **Athol Whitten** with Jim Ianelli and André Punt

UNIVERSITY of WASHINGTON

NOAA FISHERIES
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

admb
FAST, ACCURATE, STABLE OPTIMIZATION

Gmacs: the motivation, the model, and the intended applications.

- **Generalized modeling frameworks** afford many benefits
- **Gmacs will** expedite the development of new models and replacements for existing ones
- **Gmacs can** be used as part of a stream-lined stock assessment process
- **Gmacs may** also be used as an assessment research tool.

Note: Gmacs is now the name of this software. Though the name derives from 'Generalized Model for Alaska Crab Stocks' it is not intended to be used as an acronym. This reflects the fact that Gmacs may eventually be used to model other hard-to-age species. There are other potential users and developers interested in Gmacs from around the USA (NOAA NMFS) , and in Australia, New Zealand, Korea, and more.

Acknowledgements

Gmacs is a collaborative project between
University of Washington and **NOAA Fisheries**

Developers: Jim Ianelli, Athol Whitten
Principle Investigator: André Punt
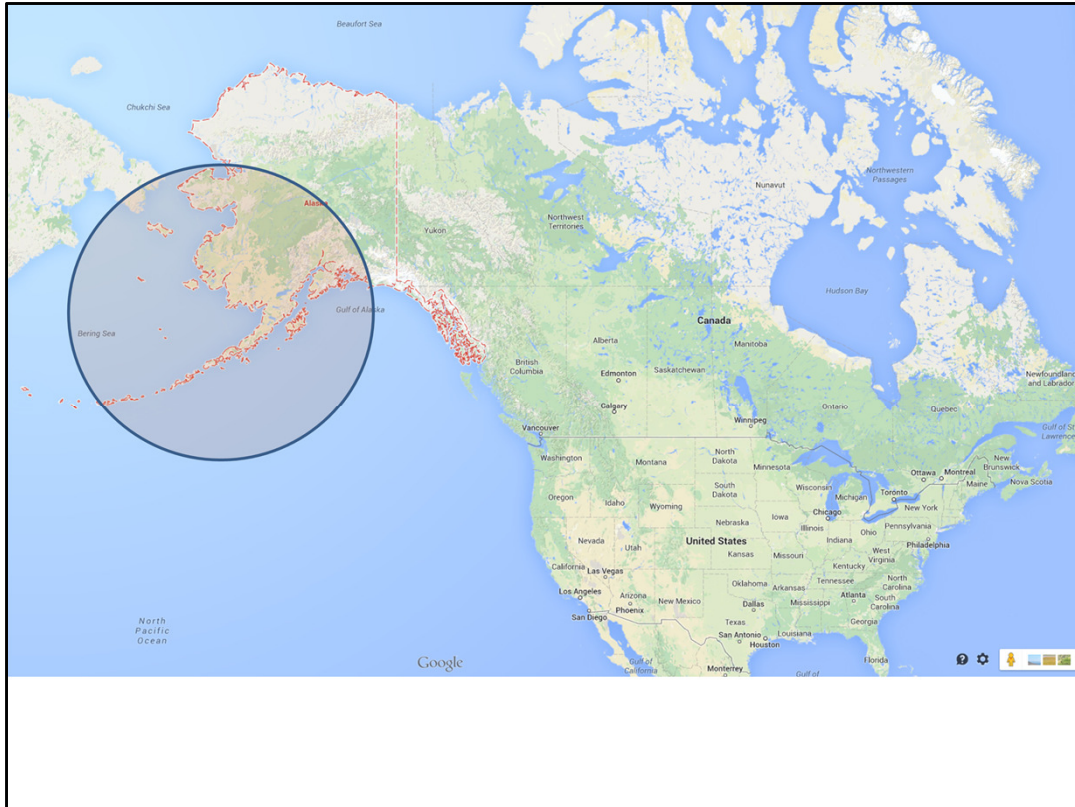**Initial Concept**: Mark Maunder

For: **North Pacific Fisheries Management Council**
and the **Bering Sea and Aleutian Islands Crab Plan Team**

Partial funding from:
Bering Sea Fisheries Research Foundation

*photo credits: NPFMC and others found online with no references.

Big Alaskan Crab Fisheries

Alaska is a big place (much bigger than Texas), with expansive marine resources. The Alaskan crab populations supports some of the worlds largest shell-fish fisheries. These fisheries are famous for their high value, and high associated risks. The sustainable management of these stocks is very important to local communities, industry, and the Alaskan and USA populations in general.

These fisheries are sometimes associated with the '**World's Deadliest Catch':** There was around **60 million pounds** of allocated catch in 2012/2013: with around 30+ million pounds of Snow Crab and around 8+ million pounds of Red King Crab actually caught.

Due to the high value of these fisheries, and their high associated risks, there is **pressure to improve the overall management of the fishery, including a desire to improve the overall assessment and management process.**

North Pacific Fisheries Management Council manages 10 crab stocks

The North Pacific Management Council (NPFMC) overseas the management of 10 Bering Sea and Aleutian Island (BSAI) Crab Stocks:
- 4 Red King Crab
- 2 Blue King Crab
- 2 Golden King Crab
- Tanner Crab and Snow Crab.

Fun fact: King Crabs are not *true* crabs; they are crab-like decapods that exist as part of their own family (*Lithodidae*), true crabs belong in another Infraorder: (*Brachyura).*

The NPFMC is one of eight regional councils established by the Magnuson-Stevens Fishery Conservation and Management Act in 1976 to manage fisheries in the 200-mile Exclusive Economic Zone around the USA.

As part of the NOAA Stock Assessment Improvement Plan: this project was formed to help improve the overall stock assessment process, especially in regard to the development of stock assessment models, for the BSAI Crab Stocks.

Crab Stock Assessment Process

Size structured assessment models

- **The lesser studied, but no less important cousin to age-structured stock assessment models**

    Punt, Huang, and Maunder (2013) Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES JMS, 70(1), 16-33.

- **Common for hard to age species:** abalone, rock lobster, prawns, other molluscs and crustaceans

There are currently size-based integrated assessment models for each of the Tier 1 species. How many?

- These are like statistical catch-at-age models but are size-structured instead
- Model fit to catch-at-size, size-frequency data, etc.
- Biological traits and fishery specifications are functions of size
- Growth is modeled via size-transition matrices (which themselves can be calculated in-part from growth curves), rather than the growth curves typical in age-based models.

Image: example of existing code.

The current stock assessment modeling situation: Existing models and model code are written and managed by individual modelers.

Code is sometimes passed from author to author, with each model's quirks and short-term solutions passed down the line, making code hard for new assessment authors to learn and maintain. Moreover, old versions of code are frequently embedded within new versions, and authors often leave redundant pieces of code in their files in order to resurrect old options, or as a basic record keeping effort. All of this makes the review process long and laborious, and sometimes near impossible for outside reviewers.

Atop of all this, each author must write and maintain complimentary code to produce plots and other diagnostics for their reports. Finally, assessment authors spend countless hours each year compiling stock assessment reports. It is hoped that the Gmacs stock assessment modeling framework will help to improve this overall stock assessment process.

Image: example of existing code.
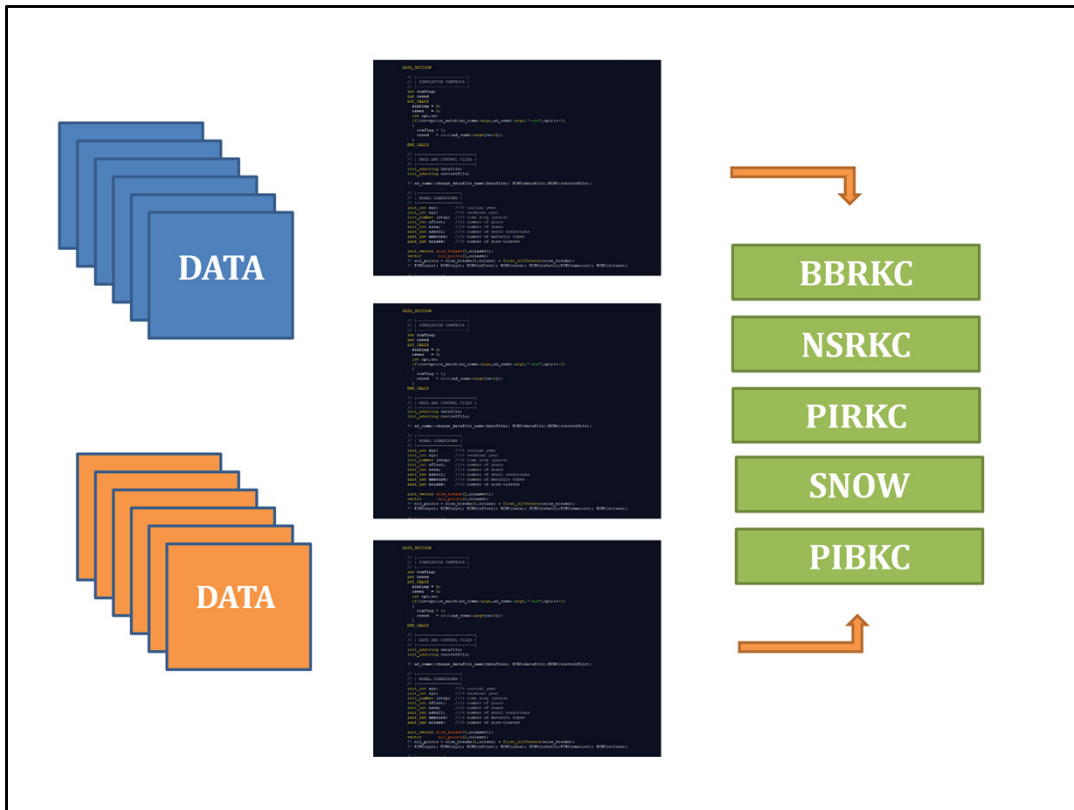
The current situation (continued): each assessment author is responsible for formatting data to be suitable for their models, and for managing his or her own modeling code: they must maintain, document, review, update, improve, and archive the details of their data processing and modelling code each year. This is a lot of work for a stock assessment author: Assessment scientists are frequently more specialised as biologists and/or applied mathematicians and statisticians, they don't like spending their time on the 'stock assessment process', but rather the models and the data. Generic Modeling Frameworks can help to improve this process.

Generalized Modeling Frameworks
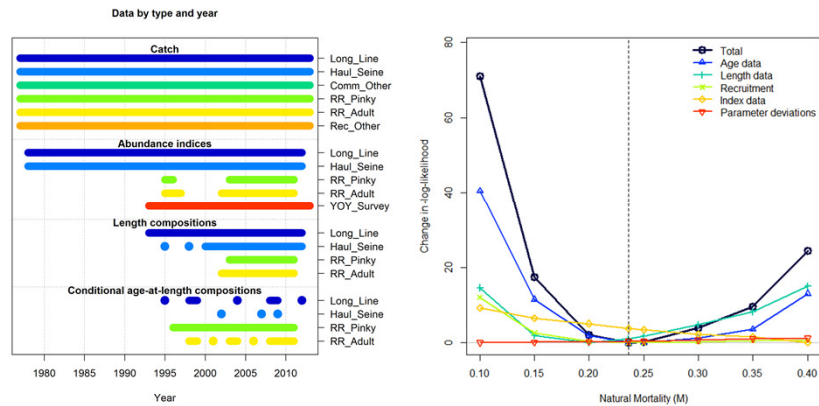
Generalized modeling frameworks

- **Many generalized modeling programs available: e.g.** MULTIFAN-CL, CASAL, Stock Synthesis

- **Gmacs** differs by being **size-structured** and **open source:** the owners of the code are stock assessment authors who contribute to the project

There are a number of successful and well-supported **generalized modeling frameworks** in existence. Gmacs follows in the footsteps of the popular Stock Synthesis platform, but differs in that it is **size-structured** and an **open-source** project - the owners of the code are stock assessment authors who contribute to the project.

Gmacs should greatly facilitate future crab stock assessment reviews, reduce errors in model formulation, expedite the development of new models for other stocks, and facilitate the transfer of models to future assessment scientists.

Assessment authors sometimes worry that using generic modeling systems will stifle their creativity, but this need not be the case. Instead, as an open-source software project, Gmacs encourages collaborators to develop their own methods and add them to the Gmacs framework as they need. Gmacs can thus provide both a useful testing ground for new methods and a framework through which such methods can be used for real assessments.

A **generalized modeling framework** permits standardised testing and reporting, which makes for **easier review.** The plots shown here are from the R package called 'r4ss' which is designed to support Stock Synthesis. By producing standardized plots and diagnostics, the r4ss package has enabled scientists, fishery managers, and industry to become well acquainted with the outputs of otherwise complex stock assessment models. An R package in support of Gmacs, called `gmr` is currently under development as part of the Gmacs project.

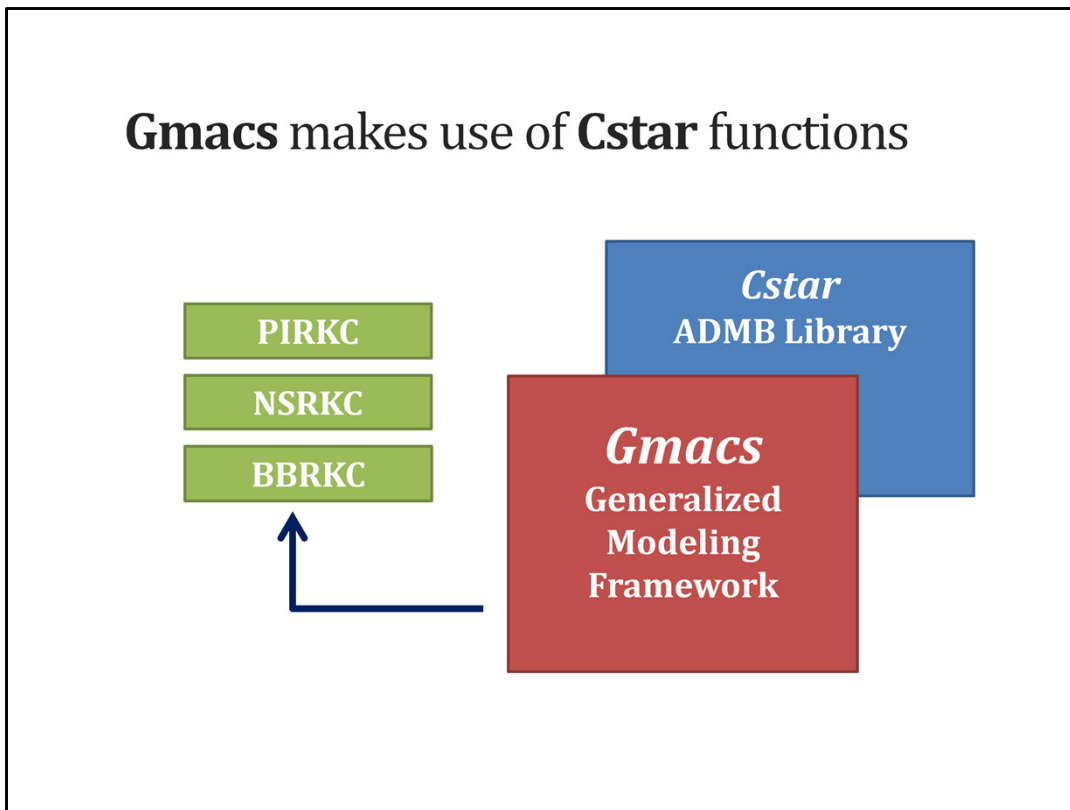Other benefits of **generalized modeling frameworks** include:
- Speeding the development of new modes, or of updating existing ones;
- Developing models that are easier to review;
- Updating standards and introducing newly accepted modeling approaches across many models and stocks in a short-time frame.

Introducing Gmacs

`https://github.com/seacode/gmacs`

Gmacs has been coded with the support of the **Cstar function library**. The Cstar function library is being developed in parallel with the Gmacs project: it is a collection of commonly used stock assessment functions that can be used in any ADMB model. The Gmacs source code does not therefore need to contain code for all of the functions which it uses, instead, these functions are inherited from the Cstar library. This minimizes the size of the Gmacs core source files (gmacs.tpl) and means it's functions can be used in other modelling projects.

Gmacs has been used to build a stock assessment model for the Bristol Bay Red King Crab stock. The Gmacs version of the BBRKC model has been tested against and compared to the established ADFG assessment model. This model serves as a testing ground for Gmacs and will be presented to the CPT on September 18th.

Gmacs general **population dynamics**

- **Gmacs implements** size-structured, integrated statistical catch-at-size assessment models

- Model can follow numbers at **time** and **size,** and by **sex, shell condition,** and **maturity**

- **Growth, molting probability** and **natural mortality** can be **time-varying** and **sex specific**

- **Time-varying parameters can be specified via** blocks, random walk processes, or as spline functions

Gmacs general **population dynamics**

- **Growth** can be specified as a simple 'parameter-per-class' option, or via a **linear growth increment** with a **gamma distribution** about the mean.

- **Maturation** and **molting probability are** specified as parametric functions of length (*can also be time-varying and sex specific*)

- **Stock recruitment estimated as yearly-deviations from a mean** *(or stock-recruitment relationship)*

Gmacs general population dynamics
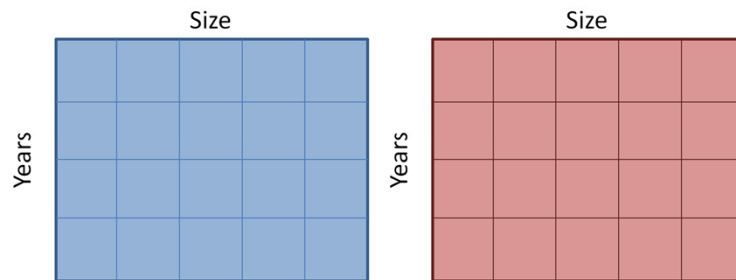
Determining suitable ways to organize Gmacs N matrix calculations:

Gmacs is designed such that the **numbers-at-size (Ns)** in a population can be dimensioned by time, sex, shell-condition, and maturity. In the simplest sense, the framework could be used to develop a single sex, single shell-type, and single maturity-type assessment model for a short time-frame. This would require only one **Ns** matrix.

However, with multiple sexes, there would need to be two **Ns matrices.**
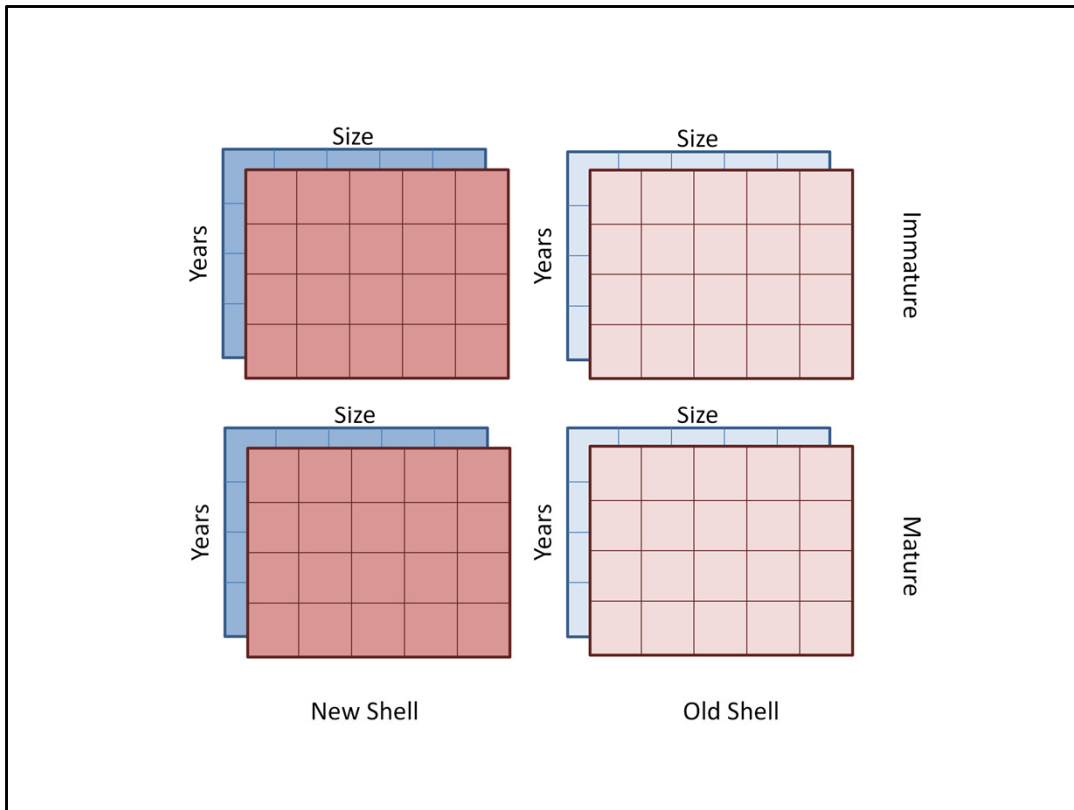
The problem quickly gets out of hand when introducing multiple shell- and maturity-types. In the case of a 2 sex, 2 shell-type (old and new) and 2 maturity-type (immature and mature) model, there is a need to track **8 Ns matrices.**
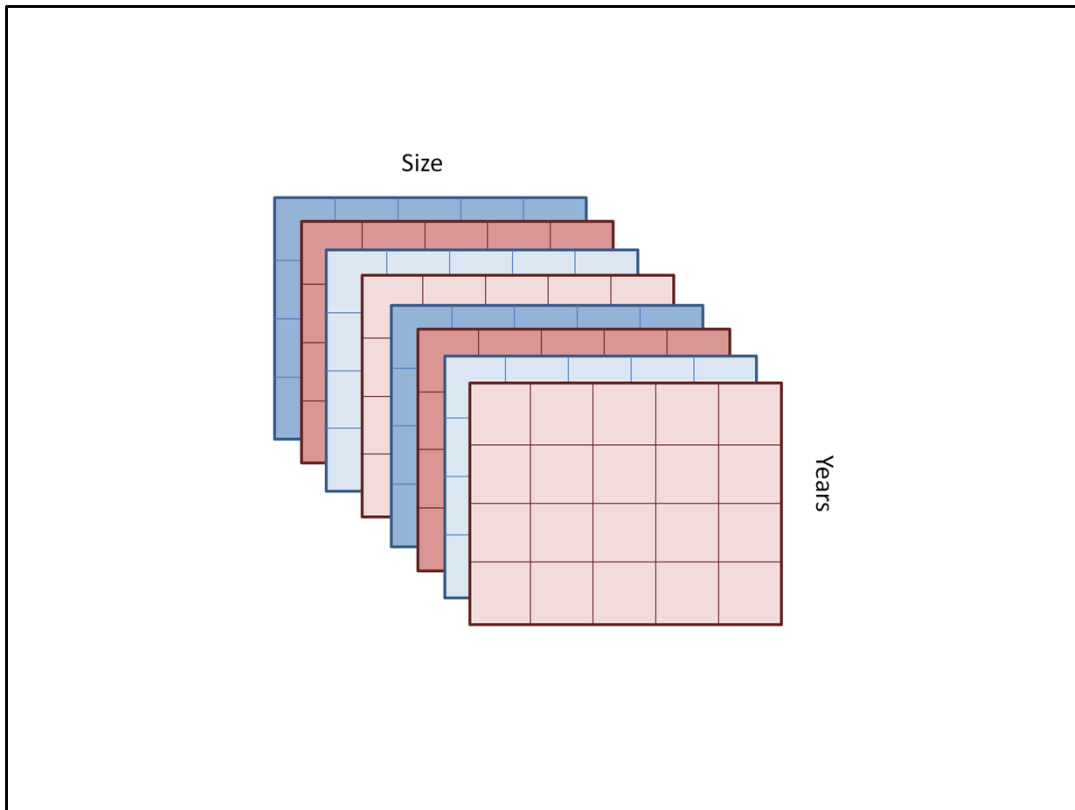
Computationally efficiency is achieved by organizing these **Ns matrices using a 3D array**, where each matrix represents the numbers at size through time, and one additional dimension. Appropriately sized arrays are specified for each model implementation by specifying the dimensions for sex, shell type, and maturity as part of the model control files.

Update this basic slide with labels and a short description.

Gmacs fits to different data types

- Model fits to **catch, discard, indices of abundance,** and **size-frequency composition data**

- Different weights can be applied to each data set

- Parameters can be fixed or estimated, and the framework allows for the specification of priors and penalties

The Gmacs modeling framework makes use of:

- Catch, discard, and by-catch data
- Fishery or survey indices of abundance
- Effort data
- Size frequency composition data
- *And soon: Tag-recapture data*

Gmacs supports a wide-range of **fishery specifications**

- **Landings** recorded from directed fisheries

- **Discards** from directed and non-target fisheries

- **Fishing mortality** estimated or directly specified

- **Selectivity** and **retention** specified as functions of length

```
DATA_SECTION

// |--------------------|
// | SIMULATION CONTROLS |
// |--------------------|
int simflag;
int rseed
LOC_CALCS
  simflag = 0;
  rseed   = 0;
  int opt,on;
  if((on=option_match(ad_comm::argc,ad_comm::argv,"-sim",opt))>-1)
  {
    simflag = 1;
    rseed   = atoi(ad_comm::argv[on+1]);
  }
END_CALCS

// |----------------------|
// | DATA AND CONTROL FILES |
// |----------------------|
init_adstring datafile;
init_adstring controlfile;

!! ad_comm::change_datafile_name(datafile); ECHO(datafile);ECHO(controlfile);

// |----------------|
// | MODEL DIMENSIONS |
// |----------------|
init_int syr;        ///> initial year
init_int nyr;        ///> terminal year
init_number jstep;   ///> time step (years)
init_int nfleet;     ///> number of gears
init_int nsex;       ///> number of sexes
init_int nshell;     ///> number of shell conditions
init_int nmature;    ///> number of maturity types
init_int nclass;     ///> number of size-classes

init_vector size_breaks(1,nclass+1);
vector       mid_points(1,nclass);
!! mid_points = size_breaks(1,nclass) + first_difference(size_breaks);
!! ECHO(syr); ECHO(nyr); ECHO(nfleet); ECHO(nsex); ECHO(nshell);ECHO(nmature); ECHO(nclass);
```

Design Features of Gmacs: Gmacs contains clean, commented code, written with flexibility and usability in mind.

Gmacs is:

• **Fully generalized** (it can handle multiple types of data, data need not be complete, nor need all types of data be available)
• **User friendly** (easy to enter data, and set control options)
• **Extendable** (readable commented code, built such that implementing new functions and features is relatively straight forward)
• **Efficient** (easy to set-up a new model, models have a fast runtime)

Data can be entered in a simple **'flat format'** and Gmacs can handle **any combination** of fleet, sex, and shell-specific data.

Furthermore, Gmacs has **simulation options** built right into the code. One of the important outcomes from the WCSAM in Boston last year was the need to perform regular simulation estimation studies on stock assessment models. Gmacs now makes this easy.
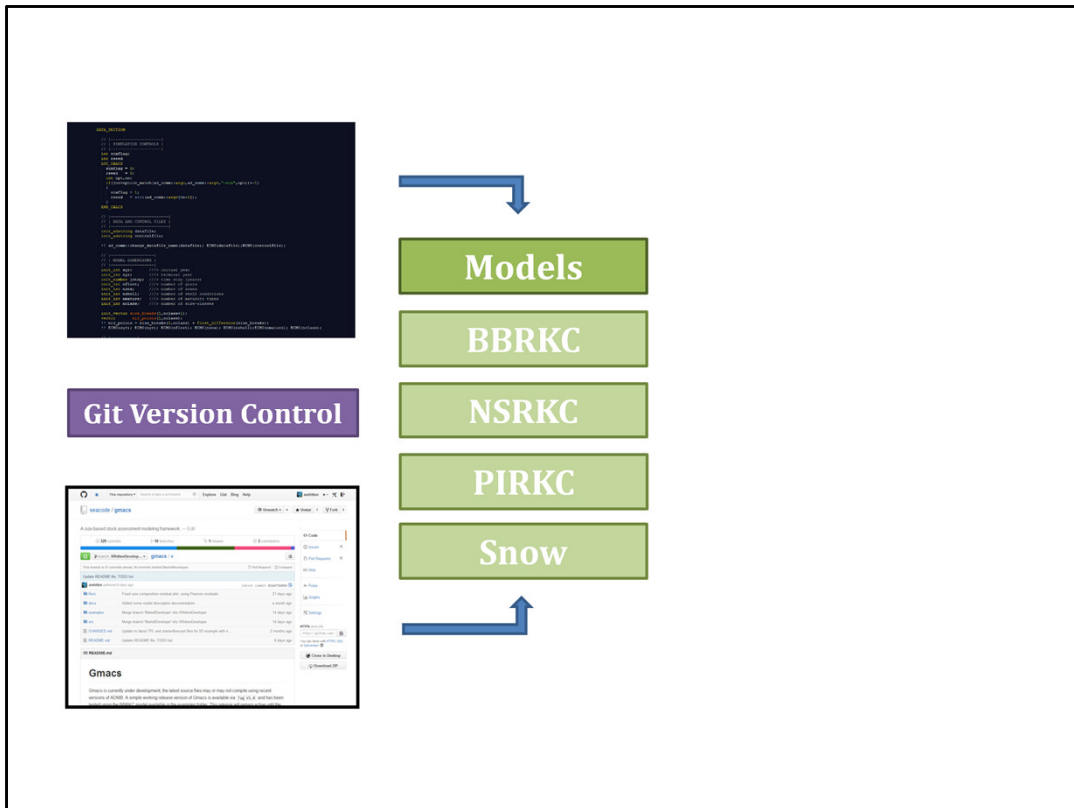
Improving the Stock Assessment Process: Stock assessment authors are continually faced with major **tradeoffs** when performing their yearly assessment tasks. Ideally the development of stock assessment models, and their associated reports, would follow the NMFS 4-T wish list:

1. **Thorough** (complete),
2. **Timely** (quick),
3. **Throughput** (many),
4. **Transparent** (understandable).

It is generally said that we can only achieve two of these in any given year, but can we get to the point where all four are possible?

Improving the stock assessment process with Gmacs: Treatment of raw data is out of the scope of this project, but Gmacs is designed to read flat-format, text-based data files which are easy to create and read. There is **one** source code file, (gmacs.tpl) which is supplemented by the Cstar library, and available online via GitHub. Different **versions** of this source code are managed by **Git Version Control Software** and each version, and it's complete history, are available online via the remote repository on GitHub. Git is a **distributed** version control program, meaning that anything that can be accessed from the remote server (in this case, Github) can also be access on a developers local computer.

Within this single modeling framework, assessment models can be developed for each stock. Modelers can utilize the features available in Gmacs to develop case-specific models, or, where necessary, add new features and incorporate them into the framework (either directly, or by contributing to the Cstar project).

The data and control files for each model can also be recorded and managed via Git and GitHub, meaning that version control for alternative data sets and model scenarios need not be a difficult task for assessment authors. Using this framework will also mean that previous years' data and models, and alternative scenarios for any given year, can be access in a matter of seconds, simply, and by anybody with access to the GitHub server.

Currently the GitHub site is an open-source repository, meaning anyone can access it.
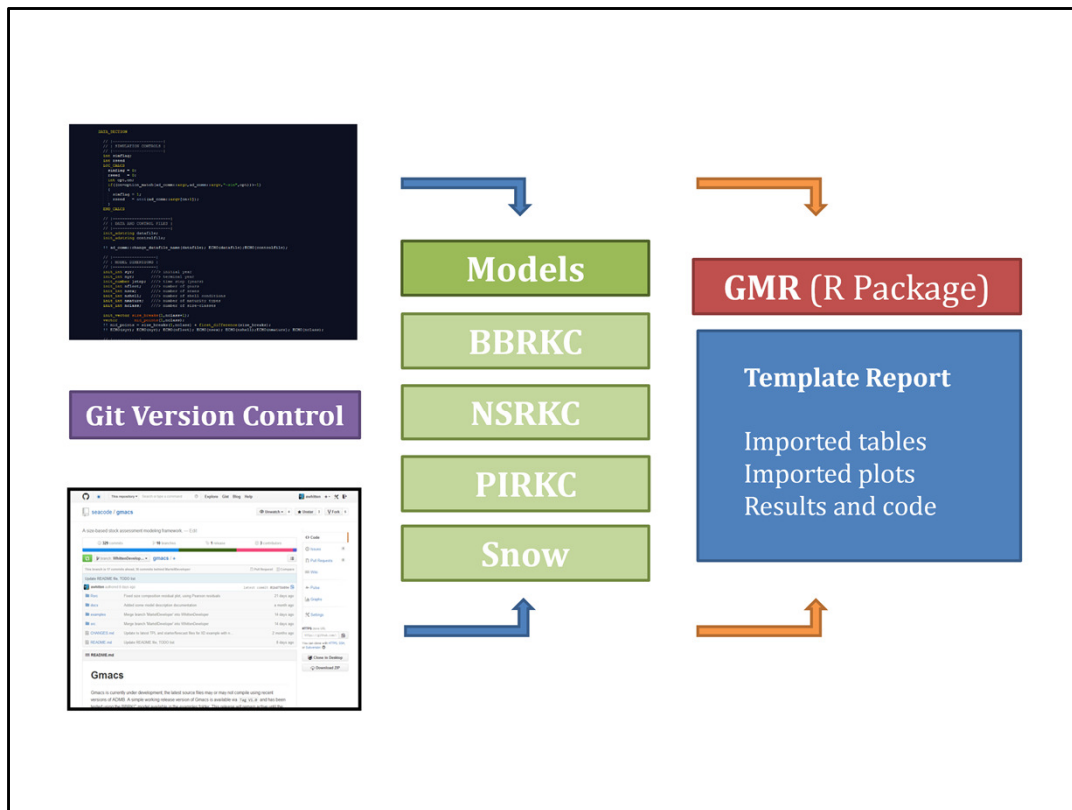
There is an R package called 'gmr' under development in support of Gmacs. It is available at https://github.com/seacode/gmr and (for the moment) focusses on plotting Gmacs outputs. The `gmr` package is also designed to be flexible and easily expandable. It follows the R package style guidelines available here: http://r-pkgs.had.co.nz/style.html and includes internal Roxygen format code-comments.

Using the 'gmr' package together with Rstudio and a simple markup language called Markdown, stock assessment reports can be produced without the laborious need to cut and paste plots and diagnostics into a WYSIWYG word-processor. With this method, an updated report can be generated whenever stock assessment models or their associated data are altered. By way of example, the Gmacs-BBRKC example assessment report has been produced using these tools. It might serve as a template for future ADFG assessment reports for BBRKC and other stocks too.

All of this reduces time spent on the **process** of doing stock assessment, and allows scientists more time to do the things that really matter: that is, modeling and data analysis! It also gives researchers more time to ask important stock-assessment research questions, and provides a system by which they can test those questions.

Collaborative Development
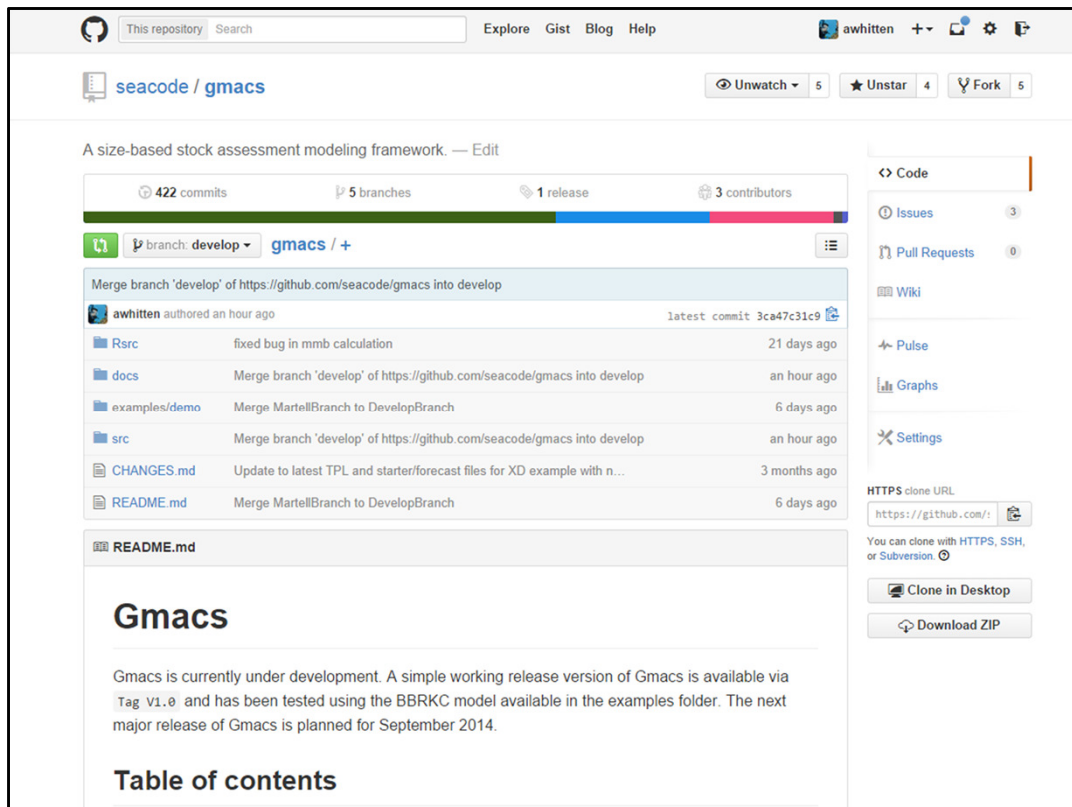
Collaborative model development is possible using Git and GitHub: users and developers can download the entire Gmacs project repositories to local machines, work locally, save versions, work on multiple models at one time (via branches), and have the entire history of the project recorded on the Github repository.

GitHub is a web-based hosting service for software development projects that use the Git revision control system. As of May 2011, GitHub was the most popular code repository site for open source projects in the world. GitHub makes it easy for collaborators to join the Cstar and Gmacs projects: Users can download code, keep up with progress, read the wiki, and report issues. Collaborators can contribute code, work on and track issues, and make their own branches for independent or collaborative development.

The Github page is now supported by a **wiki-style user-guide and developer-guide** that can be downloaded as a PDF for offline reading. The guide is still under development, but is intended to be extended by users and other developers as they become familiar with the software.

Summary

- **Generalized modeling frameworks** afford many benefits

- **Gmacs can** be used as part of a stream-lined stock assessment process

- **Gmacs will** expedite the development of new models and replacements for existing ones

- **Gmacs can** be used as an assessment research tool.

Thanks for listening

https://seacode.github.io

Dr Athol Whitten
School of Aquatic and Fishery Sciences
University of Washington | Seattle | USA

@atholwhitten | whittena@uw.edu | + 1 206 543 4270

A summary of the Gmacs project and it's supporting software, can be viewed at
https://seacode.github.io